



Bank Loan Case Study

30.07.2023

Project Description:

- The objective of this project is to conduct exploratory data analysis (EDA) on loan application data to gain insights into the patterns and characteristics associated with loan defaults.
- The aim is to gain a comprehensive understanding of the data's structure, identify factors that indicate a higher likelihood of clients facing difficulties in repaying their installments, uncover meaningful patterns, and detect potential driving factors linked to loan defaults.
- These identified factors can be utilized to make informed decisions, such as rejecting loan applications, adjusting loan amounts or loan terms, or offering loans to higher-risk applicants with increased interest rates or even denying loan applications.

- This project involves applying exploratory data analysis techniques to thoroughly examine loan application data.
- By performing EDA, the company seeks to obtain valuable insights into the risk profile of loan applicants.
- Armed with this information, the company can make data-driven decisions in managing loan portfolios and implementing risk assessment strategies.
- These insights will empower the company to take proactive measures to mitigate risks and maintain a balanced loan portfolio.

Tech - Stack Used:



- All the analysis has been performed in Microsoft Excel 2019.
- This tool is also used to create graphical representation of the results and to understand the result set better.
- To prepare the presentation I used in google slide.

Dataset Used: [application.csv](#)

Project Workflow:

Data Cleaning

Detecting Outliers

Analyze Data Imbalance

Perform Analysis

Identifying Top Correlations

- Dropped null rows starting from 50001st row till 307512th row in majority of columns.
- Further I dropped the columns having $\geq 40\%$ missing data.
- I also additional dropped columns that are irrelevant and non-related containing garbage random numbers.
- For those columns with $<40\%$ missing data are imputed with mode and median values.
- Outliers are data-points that are very unexpected and they usually deviate a lot from normal range of values.
 - I have used box plots to identify outliers for each numerical columns.
- Data imbalance can affect the accuracy of the analysis, especially for binary classification problems.
 - Understanding the data distribution is crucial for building reliable models.
 - I used pie chart to showcase the class imbalance with the distribution of target and name contract type variables.
- I performed univariate analysis to understand the distribution of individual variables, segmented univariate analysis to compare variable distributions for different scenarios, and bivariate/segmented bivariate analysis to explore relationships between variables and the target variable.
- Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.
 - I segmented the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases).
 - Identified the top correlations for each segmented data with Target=0 (loan re-payers) and Target=1 (loan defaulters).

Insights and Findings:

A. Data Cleaning:

NAME_CONTRACT_TYPE	Count of NAME_CONTRACT_TYPE	Percentage of NAME_CONTRACT_TYPE
	257512	83.74%
Cash loans	45276	14.72%
Revolving loans	4723	1.54%
Grand Total	307511	100.00%

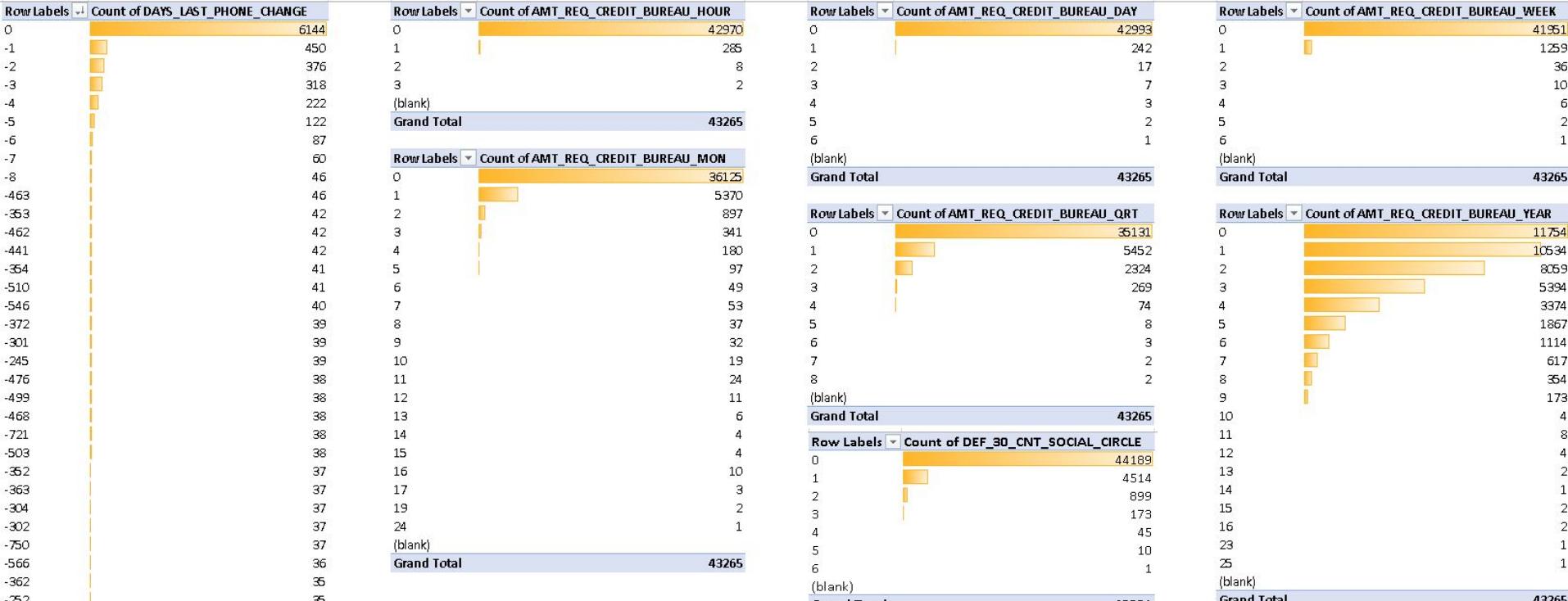
Initially, 83.74% null rows from 50001st Row till last row are removed from majority of columns.

Columns with $\geq 40\%$ i.e ≥ 20000 null cells are removed and columns with $\leq 40\%$ null cells are imputed with median and mode techniques.

Columns like [REGION_POPULATION_RELATIVE, EXT_SOURCE_2, EXT_SOURCE_3, FONDKAPREMONT_MODE, HOUSETYPE_MODE, WALLSMATERIAL_MODE, EMERGENCYSTATE_MODE and all FLAG_DOCUMENT columns] are dropped since they are irrelevant.

Count of Non-Missing values		Count of Missing Values	
Count of SK_ID_CURR	49999	SK_ID_CURR	0
Count of TARGET	49999	TARGET	0
Count of NAME_CONTRACT_TYPE	49999	NAME_CONTRACT_TYPE	0
Count of CODE_GENDER	49999	CODE_GENDER	0
Count of FLAG_OWN_REALTY	49999	FLAG_OWN_CAR	0
Count of FLAG_OWN_CAR	49999	FLAG_OWN_REALTY	0
Count of CNT_CHILDREN	49999	CNT_CHILDREN	0
Count of AMT_INCOME_TOTAL	49999	AMT_INCOME_TOTAL	0
Count of AMT_CREDIT	49999	AMT_CREDIT	0
Count of AMT_ANNUITY	49998	AMT_ANNUITY	1
Count of AMT_GOODS_PRICE	49961	AMT_GOODS_PRICE	38
Count of NAME_TYPE_SUITE	49999	NAME_TYPE_SUITE	0
Count of NAME_INCOME_TYPE	49999	NAME_INCOME_TYPE	0
Count of NAME_EDUCATION_TYPE	49999	NAME_EDUCATION_TYPE	0
Count of NAME_FAMILY_STATUS	49999	NAME_FAMILY_STATUS	0
Count of NAME_HOUSING_TYPE	49999	NAME_HOUSING_TYPE	0
Count of REGION_POPULATION_RELATIVE	49999	REGION_POPULATION_RELATIVE	0
Count of DAYS_BIRTH	49999	DAYS_BIRTH	0
Count of DAYS_REGISTRATION	49999	DAYS_EMPLOYED	0
Count of DAYS_EMPLOYED	49999	DAYS_REGISTRATION	0
Count of DAYS_ID_PUBLISH	49999	DAYS_ID_PUBLISH	0
Count of OWN_CAR_AGE	17049	OWN_CAR_AGE	32950
Count of FLAG_MOBIL	49999	FLAG_MOBIL	0
Count of FLAG_EMP_PHONE	49999	FLAG_EMP_PHONE	0
Count of FLAG_WORK_PHONE	49999	FLAG_WORK_PHONE	0
Count of FLAG_CONT_MOBILE	49999	FLAG_CONT_MOBILE	0
Count of FLAG_PHONE	49999	FLAG_PHONE	0
Count of FLAG_EMAIL	49999	FLAG_EMAIL	0
Count of OCCUPATION_TYPE	49999	OCCUPATION_TYPE	0
Count of CNT_FAM_MEMBERS	49998	CNT_FAM_MEMBERS	1
Count of REGION_RATING_CLIENT	49999	REGION_RATING_CLIENT	0
Count of REGION_RATING_CLIENT_W_CITY	49999	REGION_RATING_CLIENT_W_CITY	0
Count of WEEKDAY_APPR_PROCESS_START	49999	WEEKDAY_APPR_PROCESS_START	0
Count of HOUR_APPR_PROCESS_START	49999	HOUR_APPR_PROCESS_START	0
Count of REG_REGION_NOT_WORK_REGION	49999	REG_REGION_NOT_WORK_REGION	0
Count of REG_REGION_NOT_WORK_REGION	49999	REG_REGION_NOT_WORK_REGION	0
Count of LIVE_REGION_NOT_WORK_REGION	49999	LIVE_REGION_NOT_WORK_REGION	0
Count of REG_CITY_NOT_LIVE_CITY	49999	REG_CITY_NOT_LIVE_CITY	0
Count of REG_CITY_NOT_WORK_CITY	49999	REG_CITY_NOT_WORK_CITY	0
Count of LIVE_CITY_NOT_WORK_CITY	49999	LIVE_CITY_NOT_WORK_CITY	0
Count of ORGANIZATION_TYPE	49999	ORGANIZATION_TYPE	0
Count of EXT_SOURCE_1	21827	EXT_SOURCE_1	28172
Count of EXT_SOURCE_2	49873	EXT_SOURCE_2	126

Count of EXT_SOURCE_3	40055	EXT_SOURCE_3	9944	Count of TOTALAREA_MODE	25851	TOTALAREA_MODE	24148
Count of APARTMENTS_AVG	24614	APARTMENTS_AVG	25385	Count of WALLSMATERIAL_MODE	49999	WALLSMATERIAL_MODE	0
Count of BASEMENTAREA_AVG	20800	BASEMENTAREA_AVG	29199	Count of EMERGENCYSTATE_MODE	49999	EMERGENCYSTATE_MODE	0
Count of YEARS_BEGINEXPLOITATION_AVG	25605	YEARS_BEGINEXPLOITATION_AVG	24394	Count of OBS_30_CNT_SOCIAL_CIRCLE	49831	OBS_30_CNT_SOCIAL_CIRCLE	168
Count of YEARS_BUILD_AVG	16760	YEARS_BUILD_AVG	33239	Count of DEF_30_CNT_SOCIAL_CIRCLE	49831	DEF_30_CNT_SOCIAL_CIRCLE	168
Count of COMMONAREA_AVG	15039	COMMONAREA_AVG	34960	Count of OBS_60_CNT_SOCIAL_CIRCLE	49831	OBS_60_CNT_SOCIAL_CIRCLE	168
Count of ELEVATORS_AVG	23348	ELEVATORS_AVG	26651	Count of DEF_60_CNT_SOCIAL_CIRCLE	49831	DEF_60_CNT_SOCIAL_CIRCLE	168
Count of ENTRANCES_AVG	24804	ENTRANCES_AVG	25195	Count of DAYS_LAST_PHONE_CHANGE	49998	DAYS_LAST_PHONE_CHANGE	1
Count of FLOORSMAX_AVG	25124	FLOORSMAX_AVG	24875	Count of FLAG_DOCUMENT_2	49999	FLAG_DOCUMENT_2	0
Count of FLOORSMIN_AVG	16105	FLOORSMIN_AVG	33894	Count of FLAG_DOCUMENT_3	49999	FLAG_DOCUMENT_3	0
Count of LANDAREA_AVG	20278	LANDAREA_AVG	29721	Count of FLAG_DOCUMENT_4	49999	FLAG_DOCUMENT_4	0
Count of LIVINGAPARTMENTS_AVG	15773	LIVINGAPARTMENTS_AVG	34226	Count of FLAG_DOCUMENT_5	49999	FLAG_DOCUMENT_5	0
Count of LIVINGAREA_AVG	24862	LIVINGAREA_AVG	25137	Count of FLAG_DOCUMENT_6	49999	FLAG_DOCUMENT_6	0
Count of NONLIVINGAPARTMENTS_AVG	15285	NONLIVINGAPARTMENTS_AVG	34714	Count of FLAG_DOCUMENT_7	49999	FLAG_DOCUMENT_7	0
Count of NONLIVINGAREA_AVG	22427	NONLIVINGAREA_AVG	27572	Count of FLAG_DOCUMENT_8	49999	FLAG_DOCUMENT_8	0
Count of APARTMENTS_MODE	24614	APARTMENTS_MODE	25385	Count of FLAG_DOCUMENT_9	49999	FLAG_DOCUMENT_9	0
Count of BASEMENTAREA_MODE	20800	BASEMENTAREA_MODE	29199	Count of FLAG_DOCUMENT_10	49999	FLAG_DOCUMENT_10	0
Count of YEARS_BEGINEXPLOITATION_MODE	25605	YEARS_BEGINEXPLOITATION_MODE	24394	Count of FLAG_DOCUMENT_11	49999	FLAG_DOCUMENT_11	0
Count of YEARS_BUILD_MODE	16760	YEARS_BUILD_MODE	33239	Count of FLAG_DOCUMENT_12	49999	FLAG_DOCUMENT_12	0
Count of COMMONAREA_MODE	15039	COMMONAREA_MODE	34960	Count of FLAG_DOCUMENT_13	49999	FLAG_DOCUMENT_13	0
Count of ELEVATORS_MODE	23348	ELEVATORS_MODE	26651	Count of FLAG_DOCUMENT_14	49999	FLAG_DOCUMENT_14	0
Count of ENTRANCES_MODE	24804	ENTRANCES_MODE	25195	Count of FLAG_DOCUMENT_15	49999	FLAG_DOCUMENT_15	0
Count of FLOORSMAX_MODE	25124	FLOORSMAX_MODE	24875	Count of FLAG_DOCUMENT_16	49999	FLAG_DOCUMENT_16	0
Count of FLOORSMIN_MODE	16105	FLOORSMIN_MODE	33894	Count of FLAG_DOCUMENT_17	49999	FLAG_DOCUMENT_17	0
Count of LANDAREA_MODE	20278	LANDAREA_MODE	29721	Count of FLAG_DOCUMENT_18	49999	FLAG_DOCUMENT_18	0
Count of LIVINGAPARTMENTS_MODE	15773	LIVINGAPARTMENTS_MODE	34226	Count of FLAG_DOCUMENT_19	49999	FLAG_DOCUMENT_19	0
Count of LIVINGAREA_MODE	24862	LIVINGAREA_MODE	25137	Count of FLAG_DOCUMENT_20	49999	FLAG_DOCUMENT_20	0
Count of NONLIVINGAPARTMENTS_MODE	15285	NONLIVINGAPARTMENTS_MODE	34714	Count of FLAG_DOCUMENT_21	49999	FLAG_DOCUMENT_21	0
Count of NONLIVINGAREA_MODE	22427	NONLIVINGAREA_MODE	27572	Count of AMT_REQ_CREDIT_BUREAU_HOUR	43265	AMT_REQ_CREDIT_BUREAU_HOUR	6734
Count of APARTMENTS_MEDI	24614	APARTMENTS_MEDI	25385	Count of AMT_REQ_CREDIT_BUREAU_DAY	43265	AMT_REQ_CREDIT_BUREAU_DAY	6734
Count of BASEMENTAREA_MEDI	20800	BASEMENTAREA_MEDI	29199	Count of AMT_REQ_CREDIT_BUREAU_WEEK	43265	AMT_REQ_CREDIT_BUREAU_WEEK	6734
Count of YEARS_BEGINEXPLOITATION_MEDI	25605	YEARS_BEGINEXPLOITATION_MEDI	24394	Count of AMT_REQ_CREDIT_BUREAU_MON	43265	AMT_REQ_CREDIT_BUREAU_MON	6734
Count of YEARS_BUILD_MEDI	16760	YEARS_BUILD_MEDI	33239	Count of AMT_REQ_CREDIT_BUREAU_QRT	43265	AMT_REQ_CREDIT_BUREAU_QRT	6734
Count of COMMONAREA_MEDI	15039	COMMONAREA_MEDI	34960	Count of AMT_REQ_CREDIT_BUREAU_YEAR	43265	AMT_REQ_CREDIT_BUREAU_YEAR	6734
Count of ELEVATORS_MEDI	23348	ELEVATORS_MEDI	26651				
Count of ENTRANCES_MEDI	24804	ENTRANCES_MEDI	25195				
Count of FLOORSMAX_MEDI	25124	FLOORSMAX_MEDI	24875				
Count of FLOORSMIN_MEDI	16105	FLOORSMIN_MEDI	33894				
Count of LANDAREA_MEDI	20278	LANDAREA_MEDI	29721				
Count of LIVINGAPARTMENTS_MEDI	15773	LIVINGAPARTMENTS_MEDI	34226				
Count of LIVINGAREA_MEDI	24862	LIVINGAREA_MEDI	25137				
Count of NONLIVINGAPARTMENTS_MEDI	15285	NONLIVINGAPARTMENTS_MEDI	34714				
Count of NONLIVINGAREA_MEDI	22427	NONLIVINGAREA_MEDI	27572				
Count of FONDKAPREMONT_MODE	49999	FONDKAPREMONT_MODE	0				
Count of HOUSETYPE_MODE	49999	HOUSETYPE_MODE	0				



- For all Columns shown above are imputed with MODE value that is 0.
- AMT_REQ_CREDIT_BUREAU_YEAR column is imputed with median value 1 as there were plenty of outliers.

Row Labels	Count of OBS_60_CNT_SOCIAL_CIRCLE	Percentage of OCCUPATION_TYPE TARGET			Row Labels	Count of OBS_30_CNT_SOCIAL_CIRCLE	Row Labels	Count of DEF_60_CNT_SOCIAL_CIRCLE
		OCCUPATION_TYPE	0	1 Grand Total				
0	26699				0	26564	0	45723
1	7917				1	7907	1	3457
2	4923				2	4944	2	543
3	3337	Accountants	3.08%	0.16%	3.24%	3	3356	3
4	2291	Cleaning staff	1.34%	0.14%	1.48%	4	2318	4
5	1498	Cooking staff	1.72%	0.20%	1.93%	5	1518	5
6	1035	Core staff	8.37%	0.50%	8.87%	6	1036	(blank)
7	675	Drivers	5.41%	0.68%	6.09%	7	680	Grand Total
8	483	High skill tech staff	3.47%	0.24%	3.70%	8	510	CODE_GENDER
9	329	HR staff	0.18%	0.02%	0.20%	9	332	F
10	225	IT staff	0.15%	0.01%	0.16%	10	234	M
11	141	Laborers	16.06%	1.84%	17.90%	11	140	XNA
12	89	Low-skill Laborers	0.59%	0.12%	0.71%	12	98	Grand Total
13	61	Managers	6.49%	0.49%	6.98%	13	64	49999
14	36	Medicine staff	2.59%	0.21%	2.81%	14	36	
15	27	Private service staff	0.82%	0.07%	0.89%	15	26	
16	26	Realty agents	0.22%	0.03%	0.25%	16	24	
17	10	Sales staff	9.34%	0.98%	10.32%	17	13	
18	3	Secretaries	0.41%	0.02%	0.42%	18	5	
19	8	Security staff	2.03%	0.25%	2.28%	19	7	
20	6	Waiters/barmen staff	0.41%	0.05%	0.46%	20	5	
21	4	Grand Total	91.95%	8.05%	100.00%	21	4	
22	3	Percentage of NAME_TYPE_SUITE TARGET	0.37%	0.01%	0.38%	22	5	
23	1	NAME_TYPE_SUITE	0	1 Grand Total		23	1	
24	1	Children	0.99%	0.09%	1.08%	24	1	
25	10873	Family	12.10%	1.00%	13.10%	25	1	
26	25807	Group of people	0.07%	0.00%	0.07%	26	1	
27	8635	Other_A	0.25%	0.02%	0.27%	27	1	
28	4000	Other_B	0.46%	0.06%	0.52%	28	1	
(blank)	592	Spouse, partner	3.41%	0.29%	3.70%	(blank)		
Grand Total	68	Unaccompanied	74.29%	6.58%	80.87%	Grand Total	49831	
6	12	Grand Total	91.95%	8.05%	100.00%			
7	2							
8	1							
9								
10								
11								
12								
13								
(blank)								

- These columns with BLANK cells are imputed with MODE values holding most percentage of values in each columns.
- For CODE_GENDER column, 'XNA' values replaced with 'F' values.

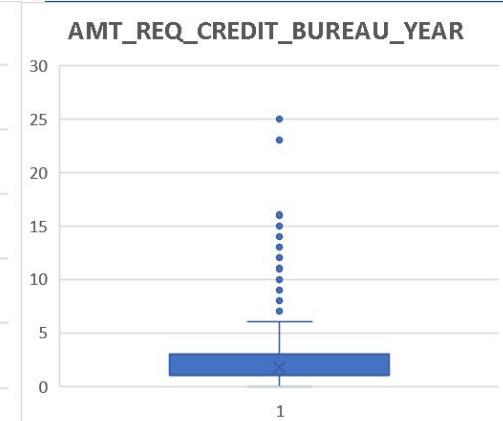
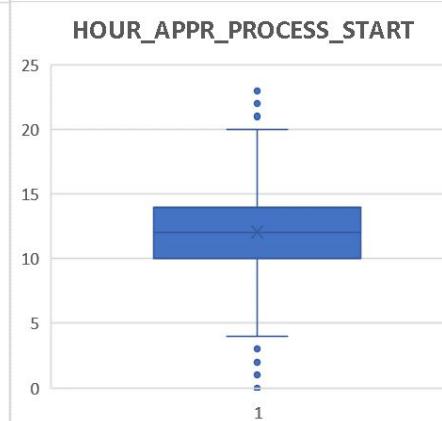
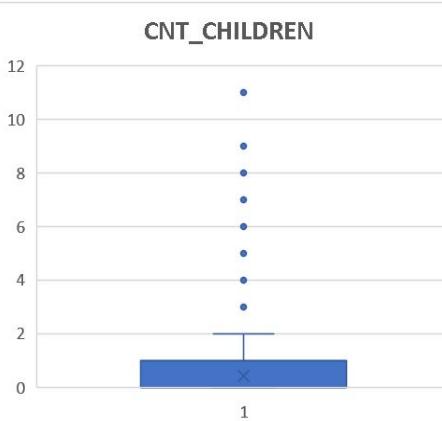
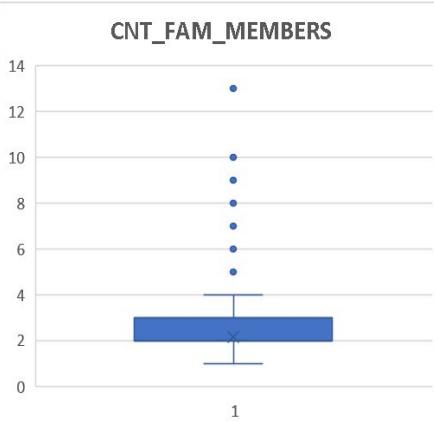
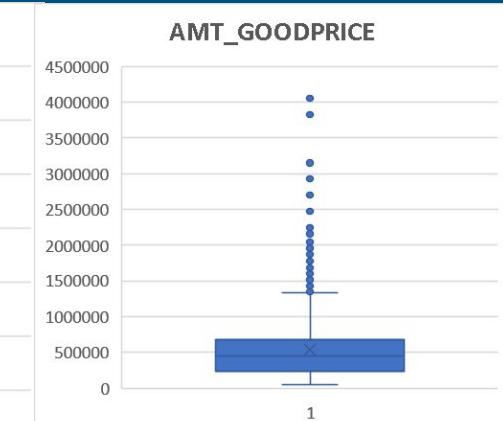
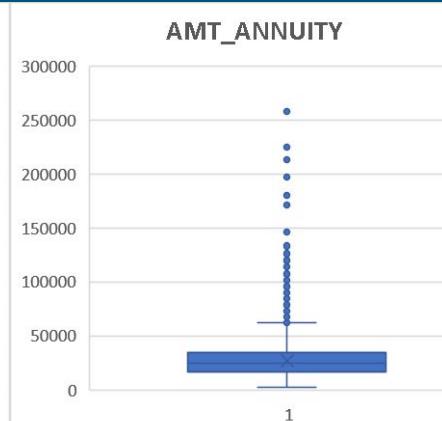
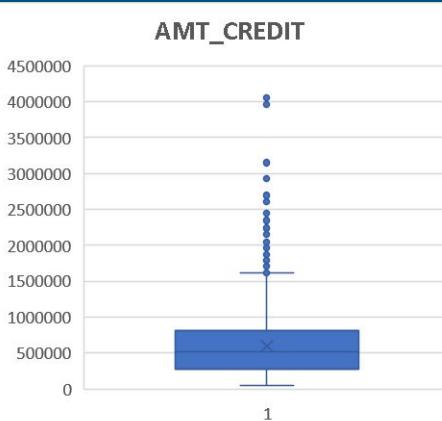
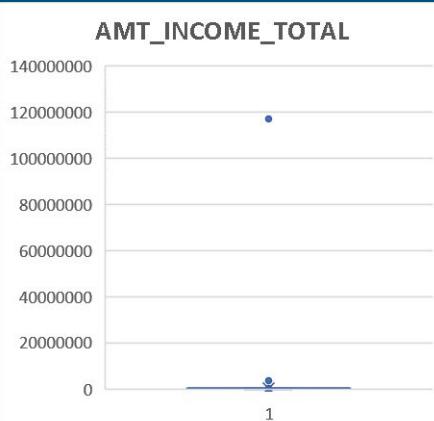
fx

=ABS(U2/365)

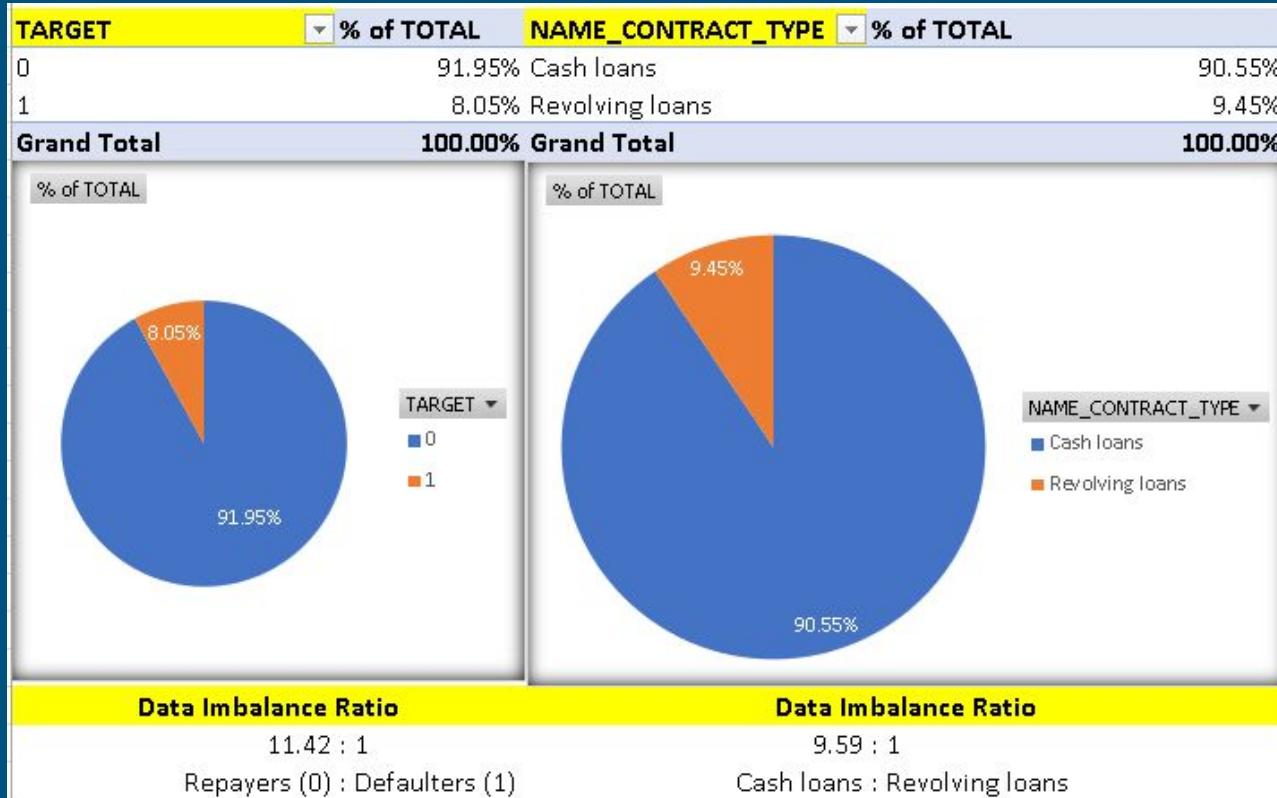
Q	R	S	T	U	V
DAY'S_BIRTH	DAY'S_BIRTH(YEARS)	DAY'S_EMPLOYED	DAY'S_EMPLOYED(YEARS)	DAY'S_REGISTRATION	DAY'S_REGISTRATION(YEARS)
-9461	26	-637	2	-3648	10
-16765	46	-1188	3	-1186	3
-19046	52	-225	1	-4260	12
-19005	52	-3039	8	-9833	27
-19932	55	-3038	8	-4311	12
-16941	46	-1588	4	-4970	14
-13778	38	-3130	9	-1213	3
-18850	52	-449	1	-4597	13

- I also added these three columns to convert these day columns into year columns to make the analysis more simpler to understand and also converted the negative value to positive value using Excel's absolute function by dividing the day columns by 365.

B. Identifying Outliers in the dataset:

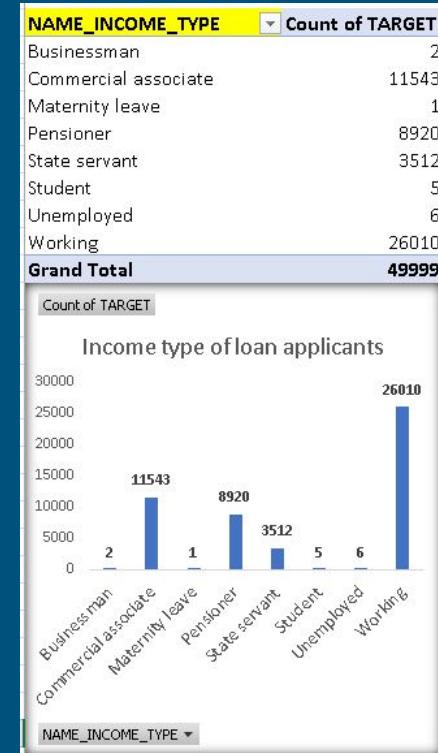
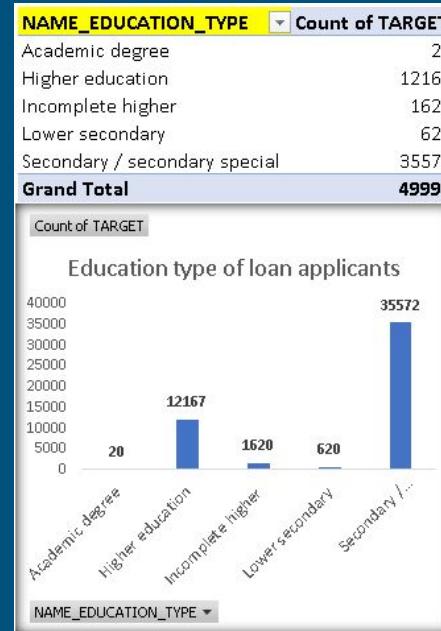
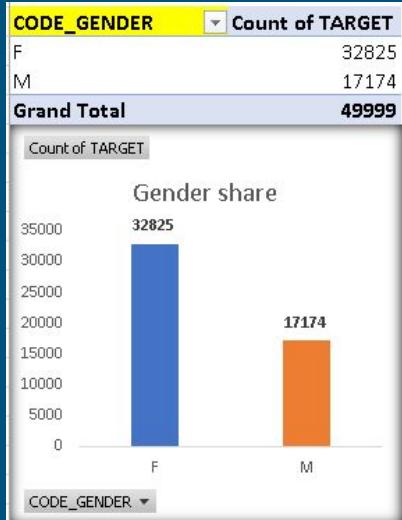


C. Data Imbalance:



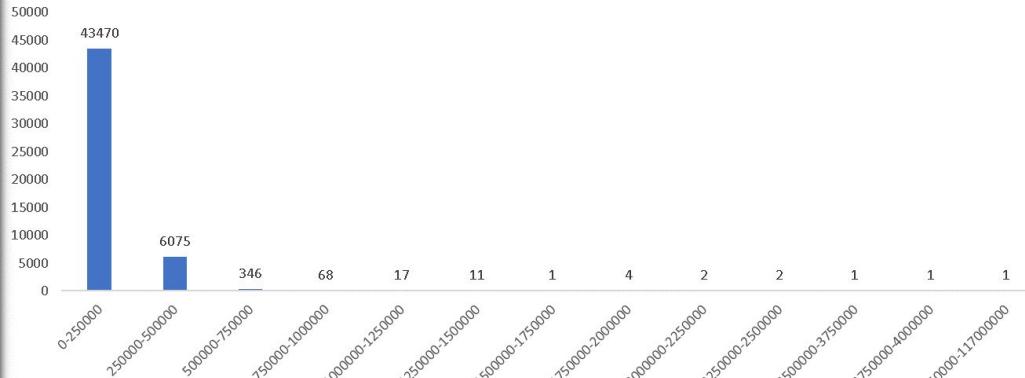
Data Analysis:

A. Univariate Analysis:



Count of TARGET

Total loan Applicants per income range



AMT_INCOME_TOTAL ▾

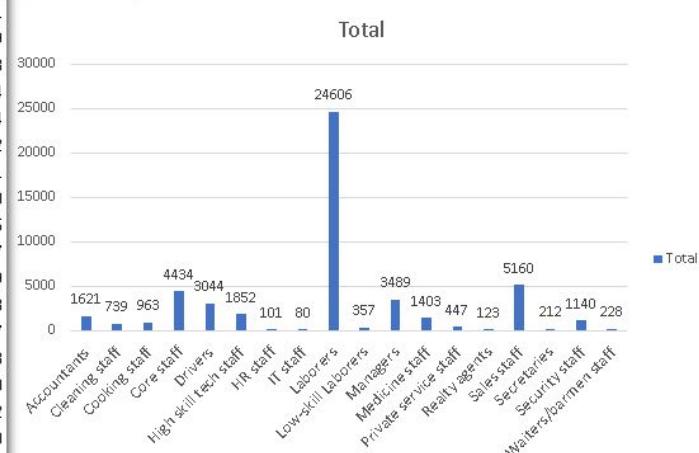
OCCUPATION_TYPE ▾ Count of TARGET

Count of TARGET

Accountants	1621
Cleaning staff	739
Cooking staff	963
Core staff	4434
Drivers	3044
High skill tech staff	1852
HR staff	101
IT staff	80
Laborers	24606
Low-skill Laborers	357
Managers	3489
Medicine staff	1403
Private service staff	447
Realty agents	123
Sales staff	5160
Secretaries	212
Security staff	1140
Waiters/barmen staff	228
Grand Total	49999

Count of TARGET

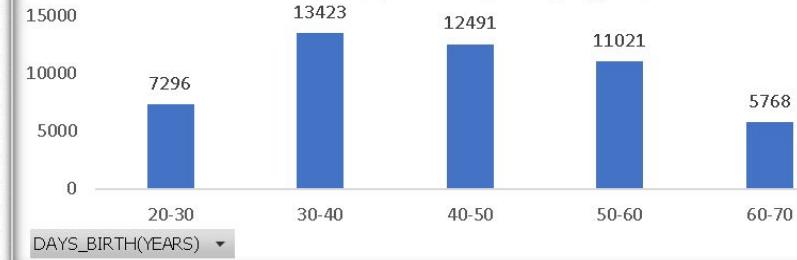
Total



OCCUPATION_TYPE ▾

Count of TARGET

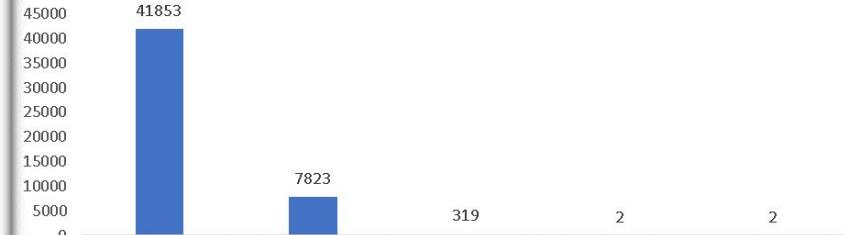
Total loan applicants per age-group



DAYS_BIRTH(YEARS) ▾

Count of TARGET

total loan applicants per Credit range

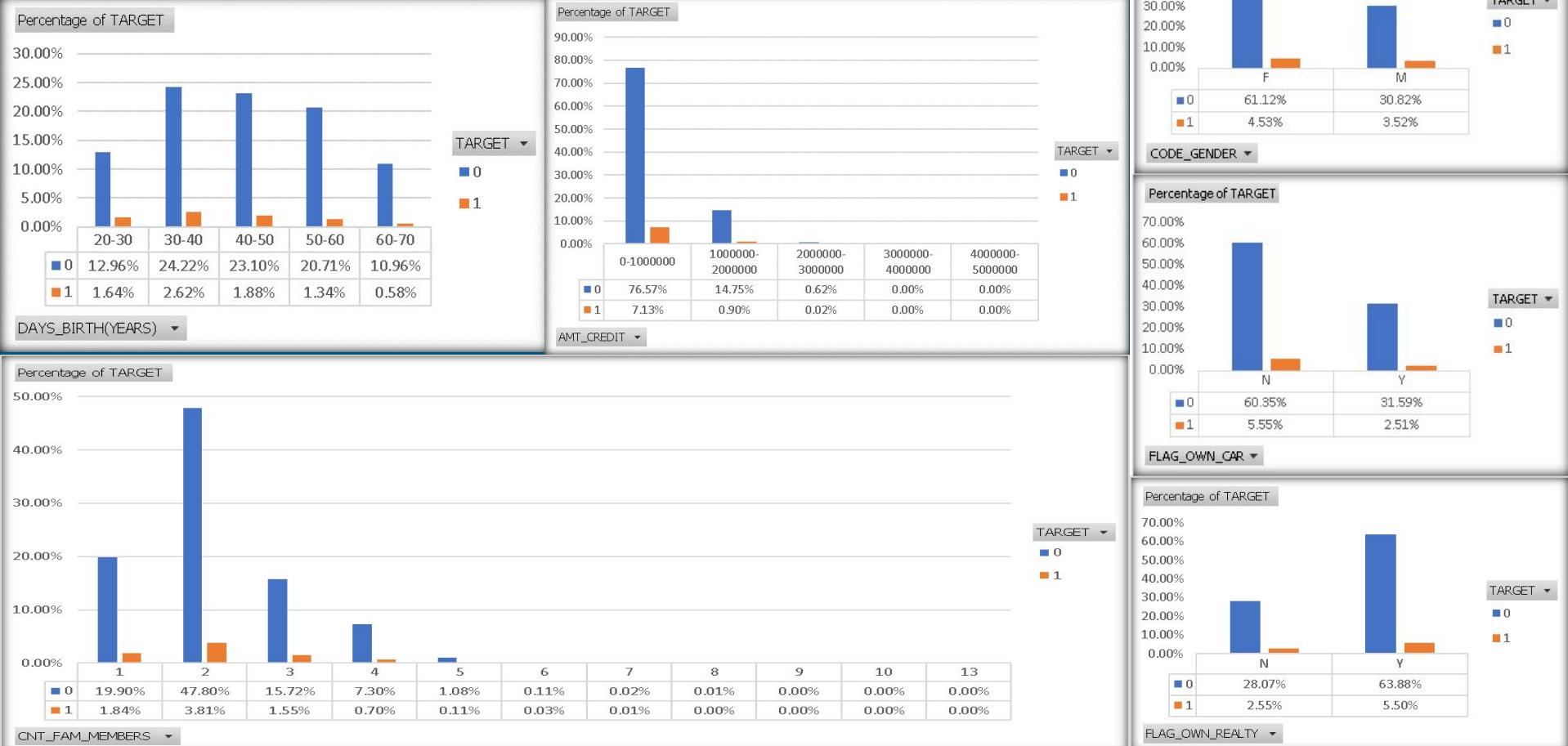


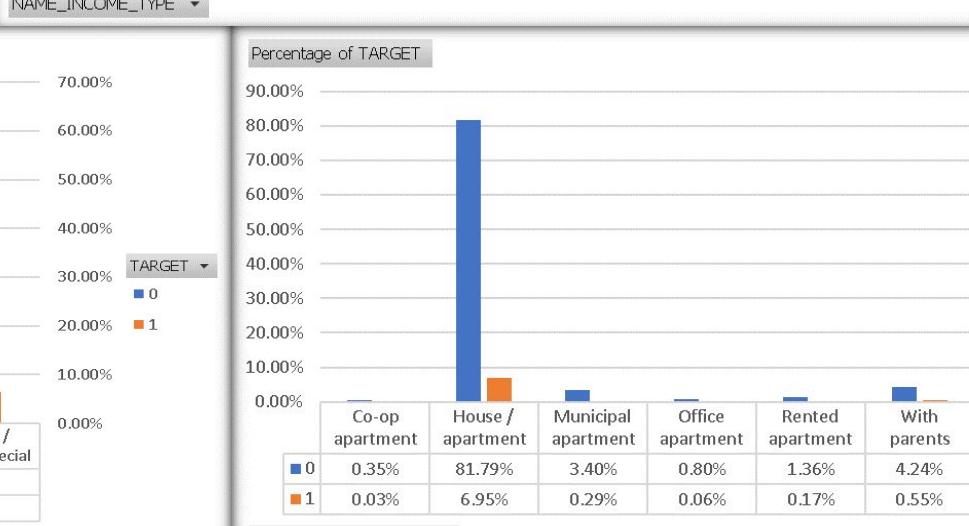
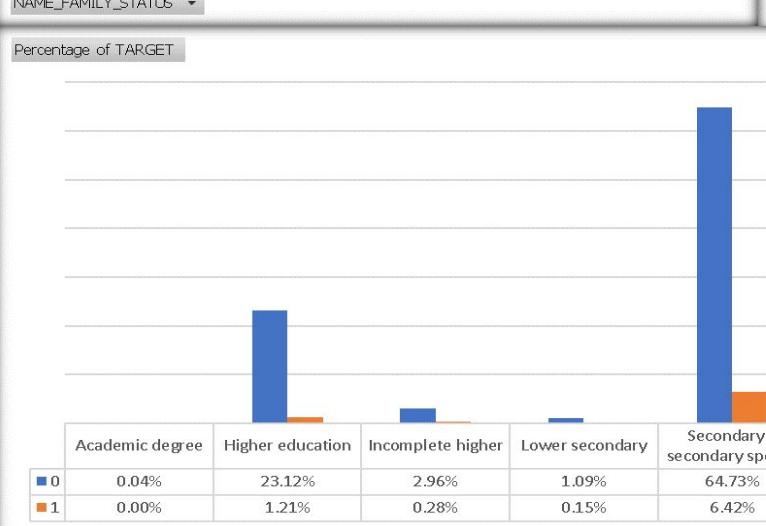
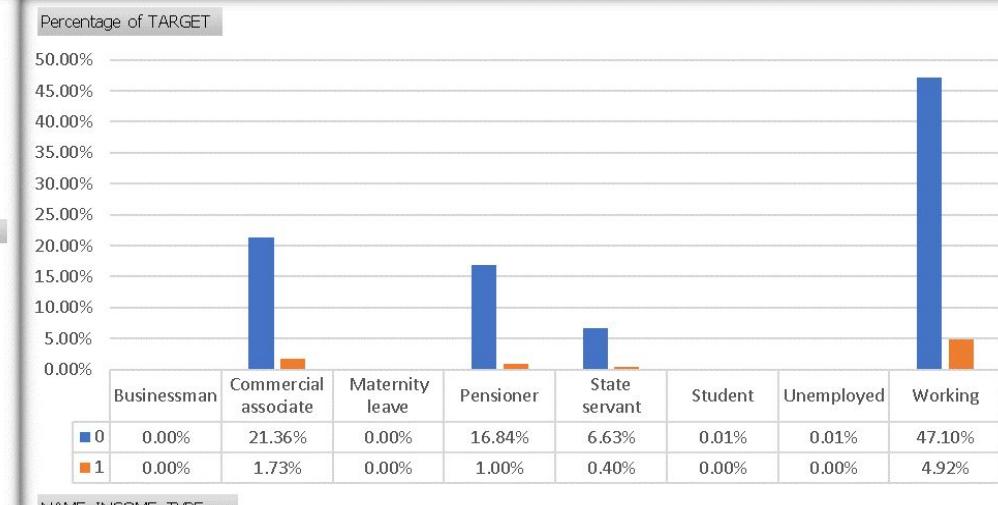
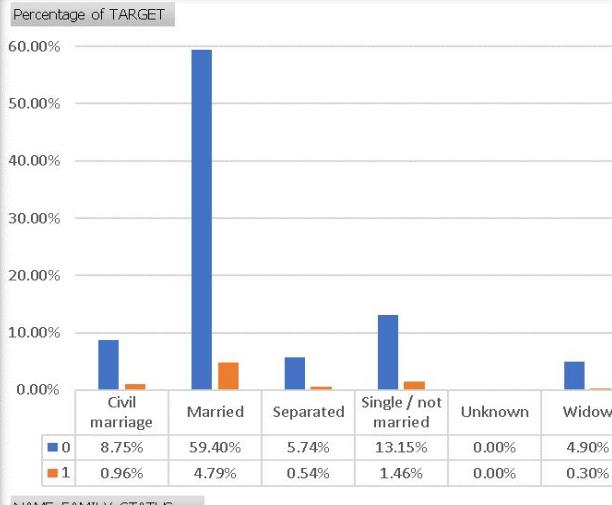
AMT_CREDIT ▾

Findings from Univariate Analysis:

- Majority of loan applicants are **female** applicants.
- Most of the people who are **married** are the major loan loan applicants.
- Majority of loan applicants are **secondary education passouts**.
- Those who are in **working** profession are the major loan applicants.
- People whose income is in the range of **0-250000** hold the majority share in loan applications.
- Those who fall under **Laborer** type of occupation tend to apply for more loans.
- People whose age comes under bracket from **30-50** are the major loan applicants.
- People who apply loan for credit amount range of **[0 - 100000]** hold the majority share in loan applications.

B. Segmented Univariate Analysis:



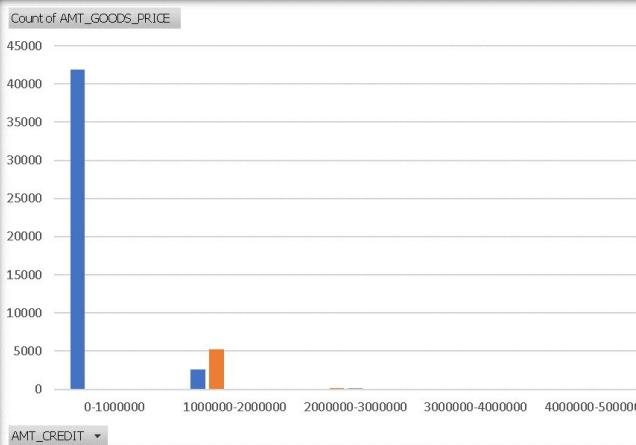


Findings from Segmented Univariate Analysis:

- Most of the loan re-payers and loan defaulters are people who aged between **30-50**, and the number of loan re-payers are more than the number of loan defaulters in all age groups.
- Credit amount demand for loan application is maximum for the range between **[0 - 1000000]** in case of both loan re-payers and loan defaulters and the demand gradually decreases for higher credit amount ranges.
- The loan demand is maximum for the client who is having **2** family members in case of both loan re-payers and loan defaulters and further it gradually **decreases** for client having higher number of family members.
- **Females** dominate **Males** in terms of both loan re-paying and loan defaulting and the loan re-payers are more than loan defaulters.
- Loan applicants who **don't own a car** hold most proportion in terms of both loan re-payment and loan defaults and the loan re-payers are than loan defaulters.

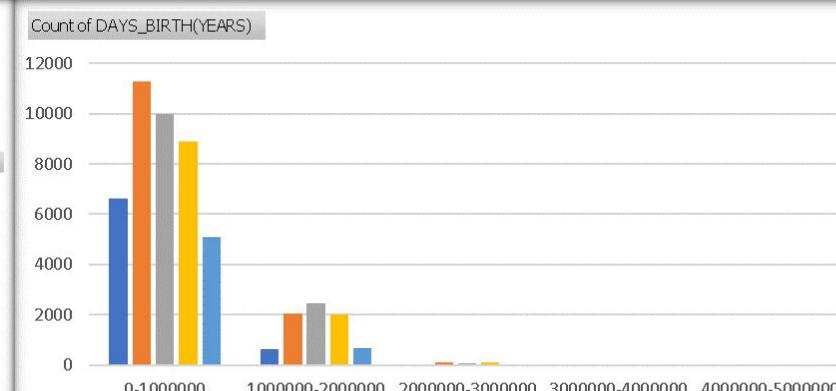
- Loan applicants who **own a realty/property** hold most proportion in terms of both loan re-payment and loan defaults and the loan re-payers are than loan defaulters.
- Most of the loan re-payers and loan defaulters are **married** people and the number of loan re-payers are more than the number of loan defaulters.
- **Secondary education pass-out** candidates are the ones who are repay the loan the most and also tends to default the most, but the number of loan re-payers are more than the number of loan defaulters.
- Most of the **working** professionals among all type of professions tend to pay the loan on time, also there are more loan defaulters among all type of professions but the number of loan re-payers are more than the number of loan defaulters.
- Most of the loan applicants who **own the House or apartment** pay the loan on time and also get defaulted the most in terms of all Housing types.

C. Bivariate Analysis:



AMT_GOODS_PRICE ▾

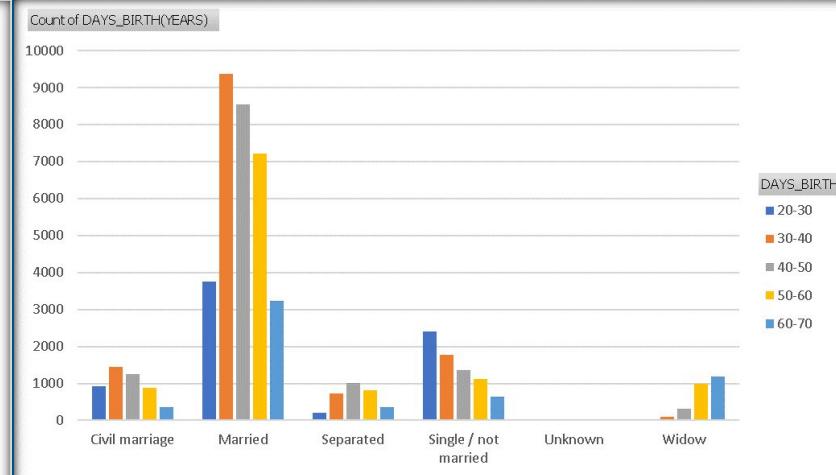
- 0-999999
- 1000000-1999999
- 2000000-2999999
- 3000000-3999999
- 4000000-4999999



DAYS_BIRTH(YEARS) ▾

- 20-30
- 30-40
- 40-50
- 50-60
- 60-70

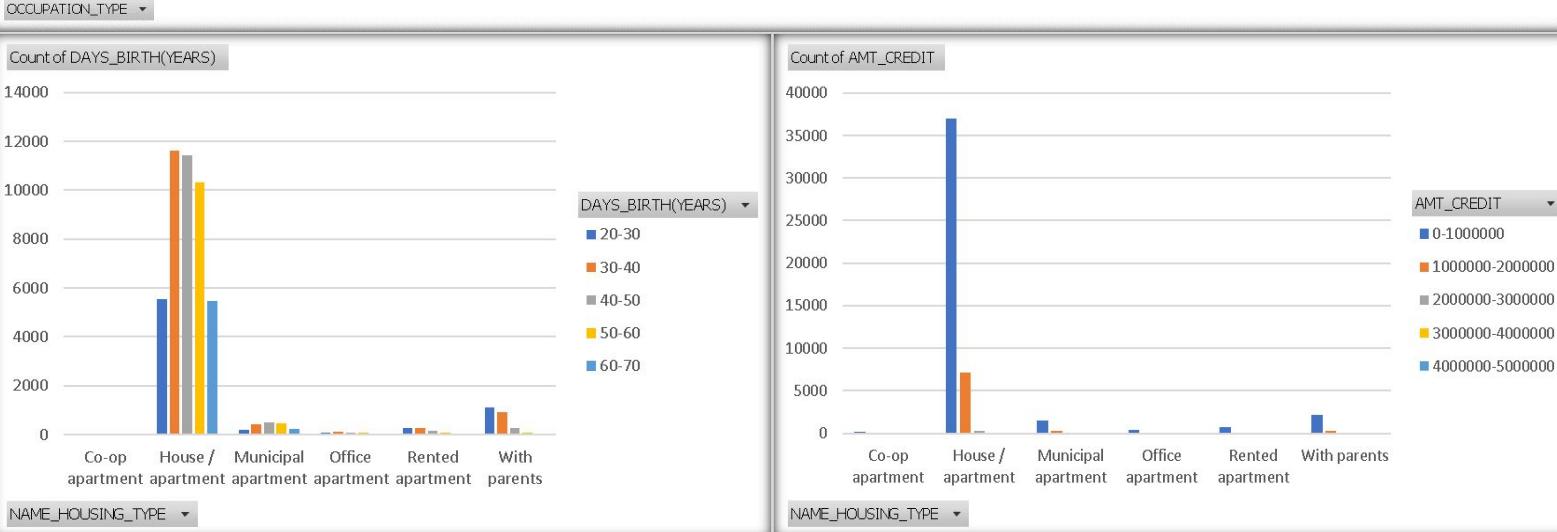
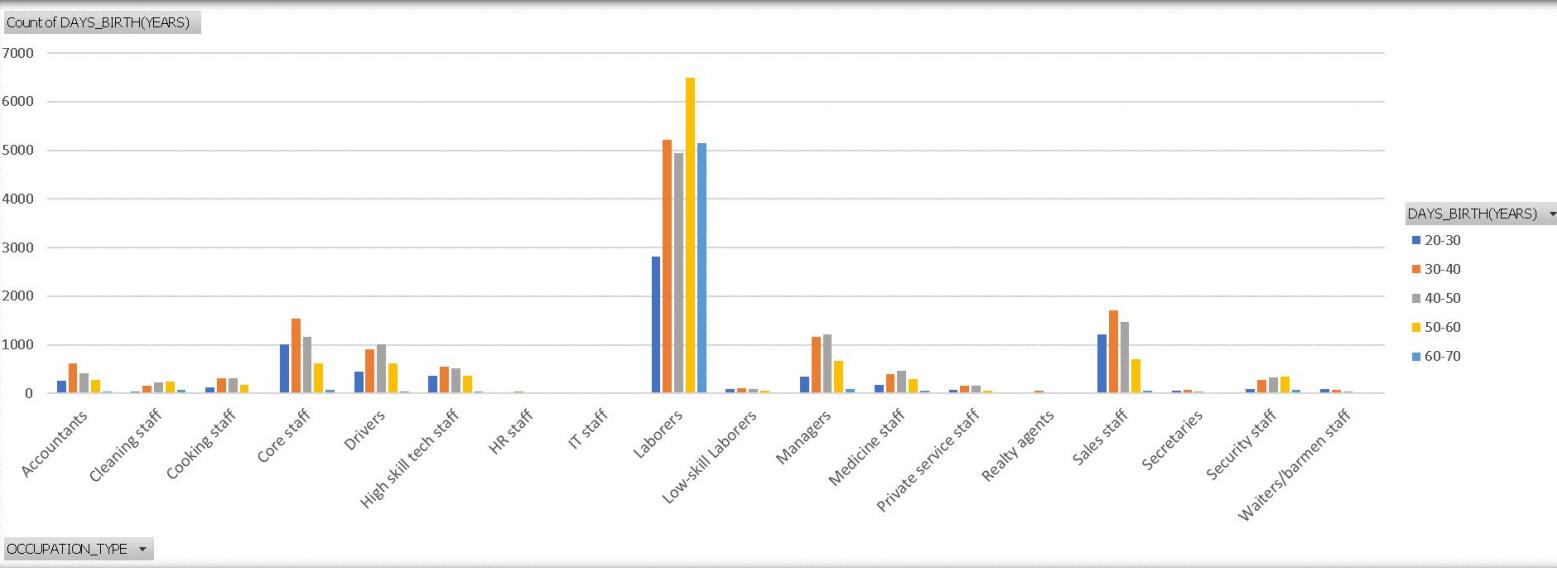
AMT_CREDIT ▾

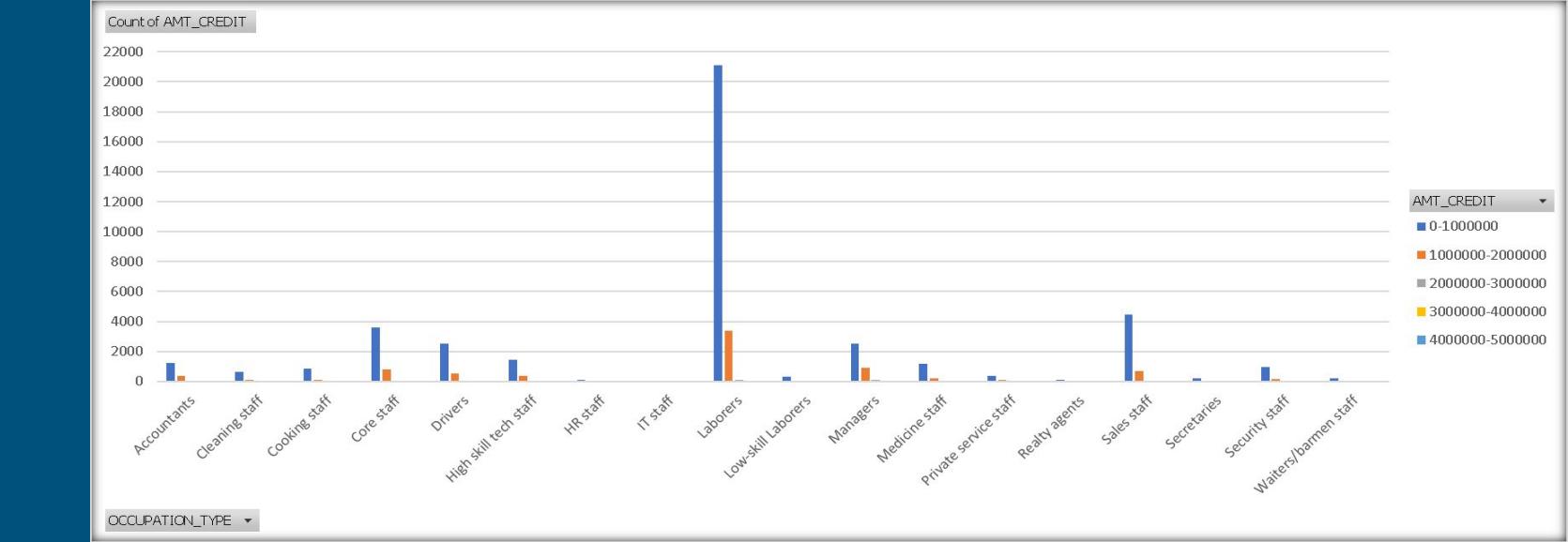
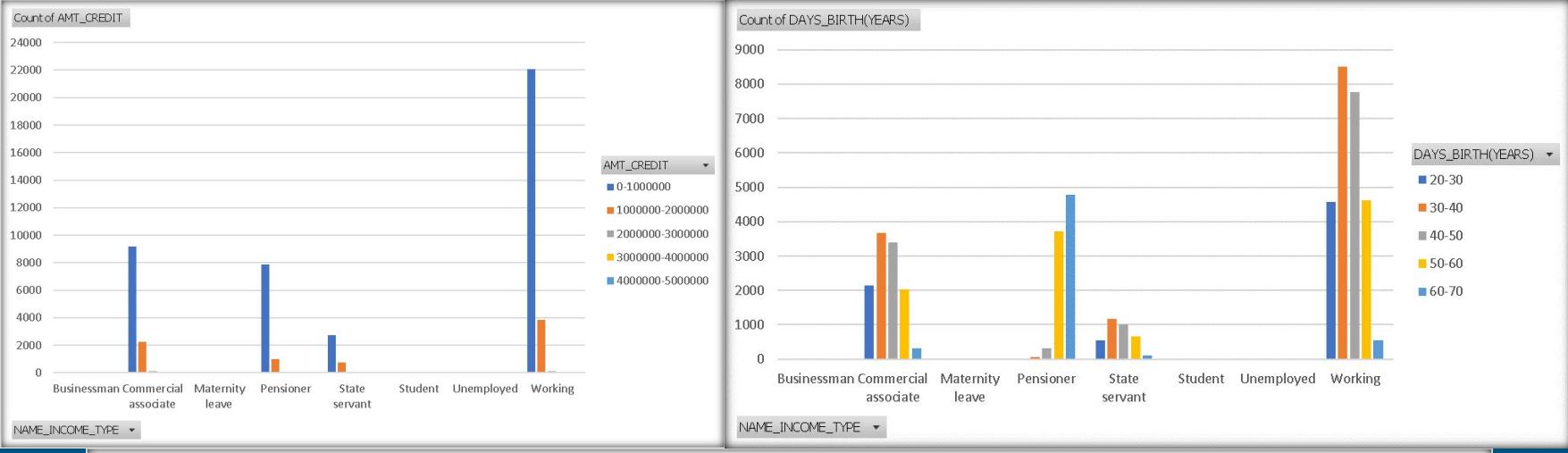


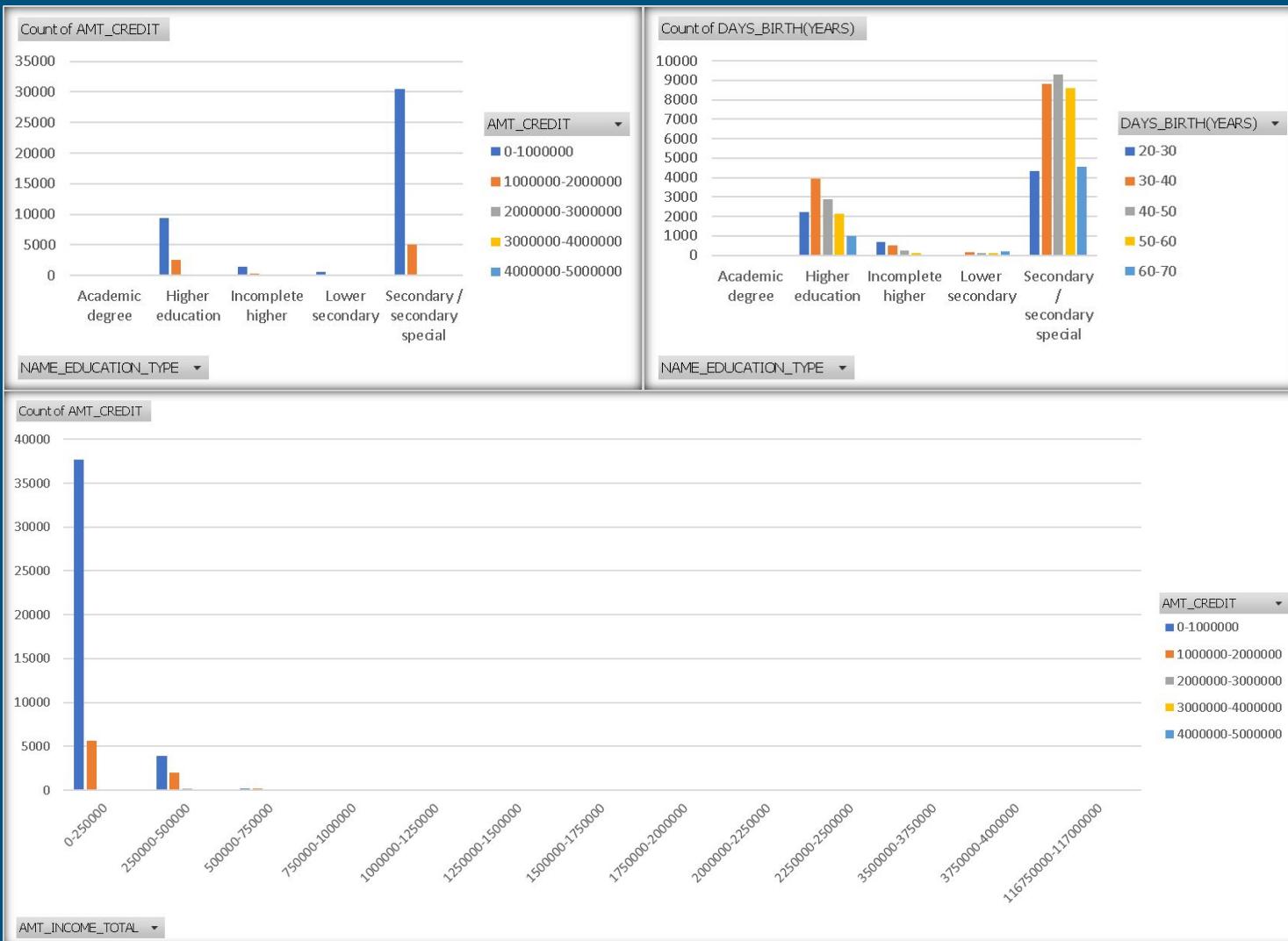
DAYS_BIRTH(YEARS) ▾

- 20-30
- 30-40
- 40-50
- 50-60
- 60-70

NAME_FAMILY_STATUS ▾





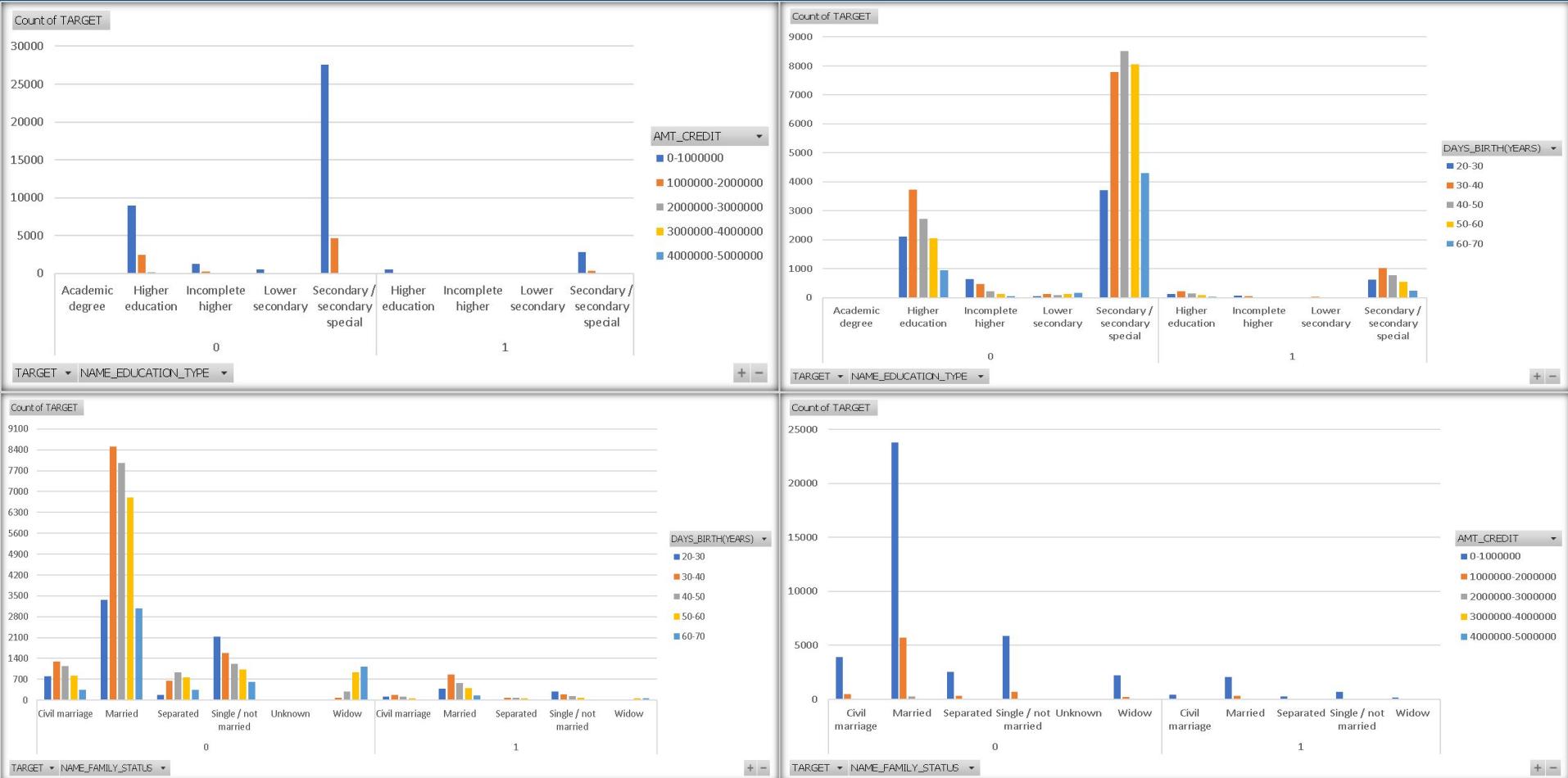


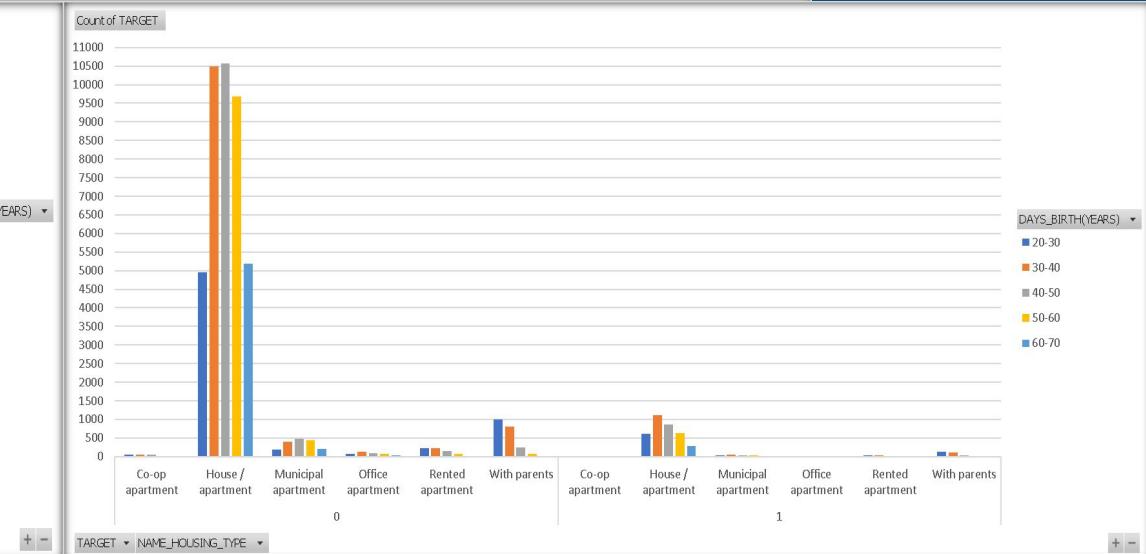
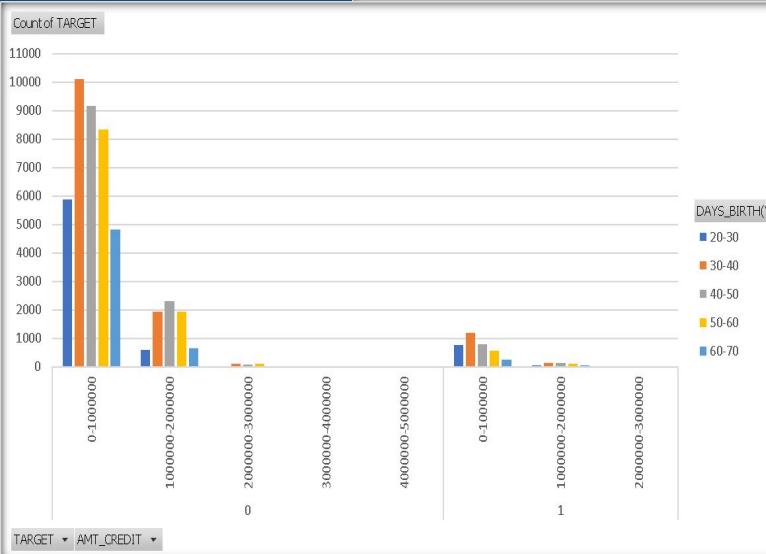
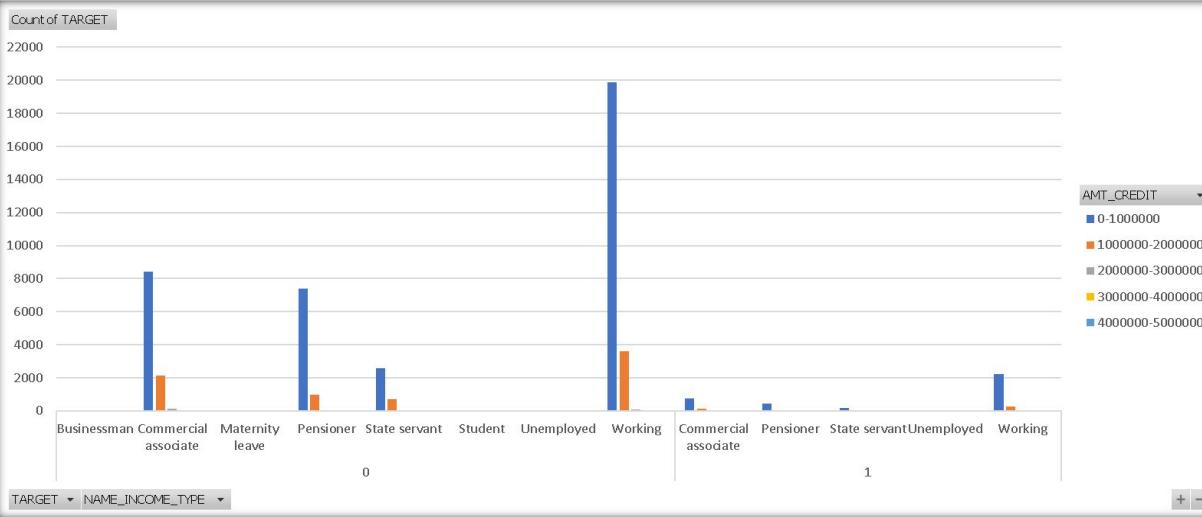
Findings from Bivariate Analysis:

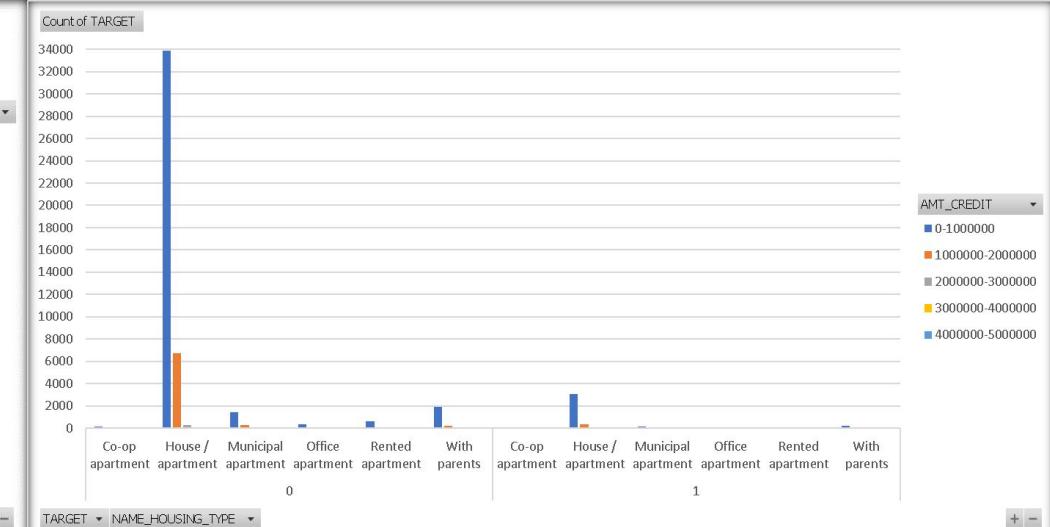
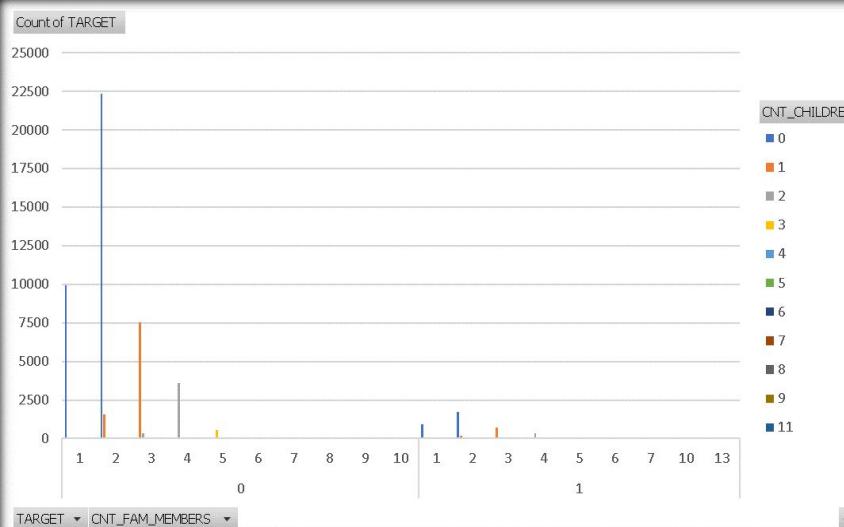
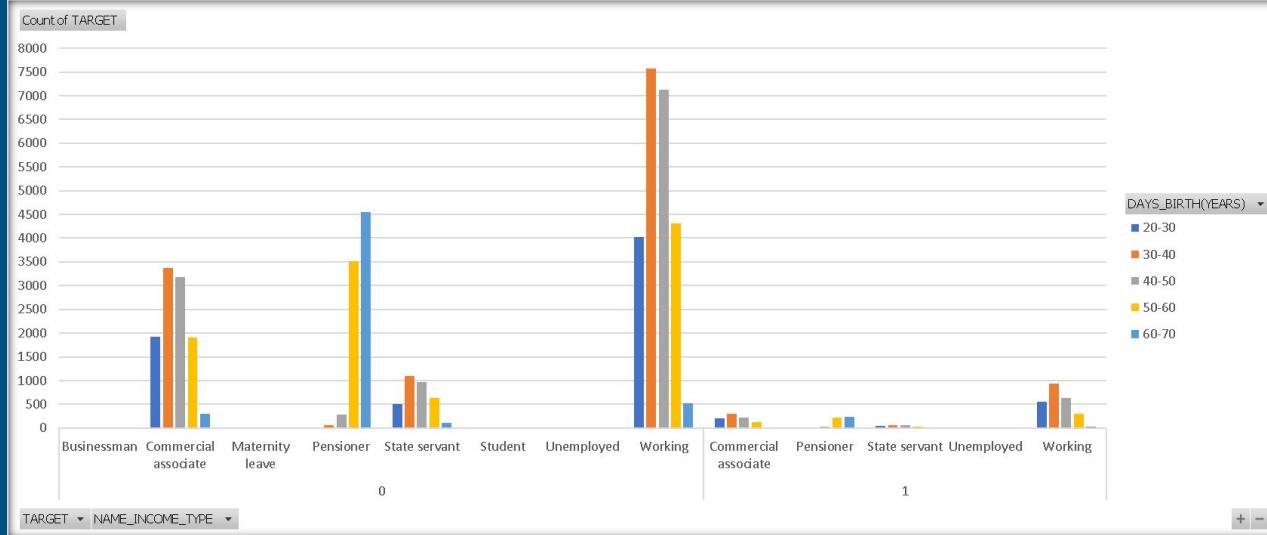
- Most of the consumers goods loans in the range of [0-999999] take the loan in the credit amount range of [0-1000000].
- Loan applicants aged between 30-50 take the loans the most that are in the credit amount range of [0-1000000].
- Most of the loan applicants among all **income types** who are **working** professionals are aged in between 30-50 years and take the loan having credit amount range of [0-1000000].
- Most of the loan applicants among all **family status** who are **married** are aged in between 30-50 years and take the loan having credit amount range of [0-1000000].
- Most of the loan applicants whose **occupation type** are **Laborers** are aged in between 30-50 years and take the loan having credit amount range of [0-1000000].
- Most of the loan applicants who own a **house or apartment** among all **housing types** are aged between 30-50 years and take the loan having credit amount range of [0-1000000].

- Most of the loan applicants who have pursued till **secondary education** among all **education types** have taken loans maximum at the credit amount range of [**0-1000000**] and most of them are aged between **40-50** years.
- Most of the loan applicants having **income range** from [**0 - 250000**] take loans more having credit amount range of [**0-1000000**].

D. Segmented Bivariate Analysis:







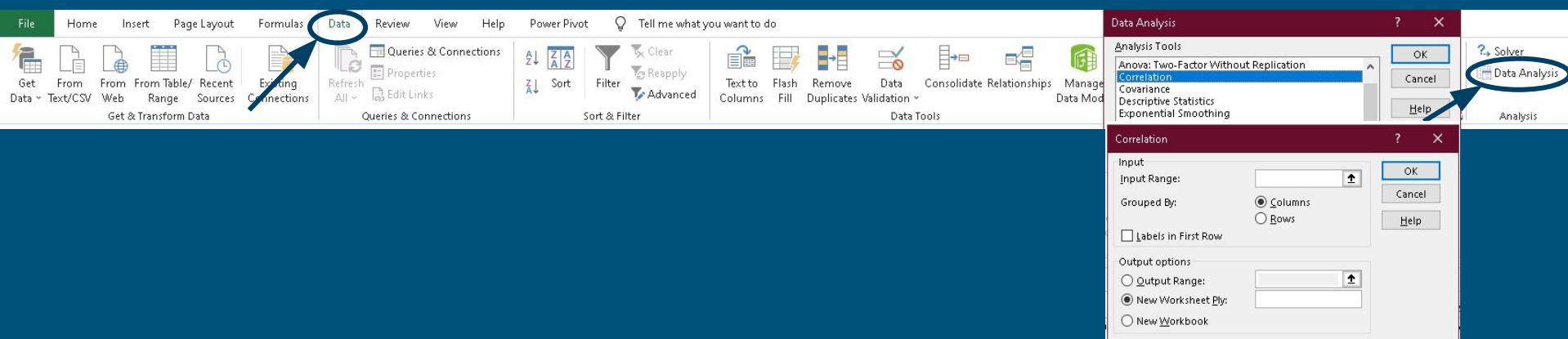
Findings from Segmented Bivariate Analysis:

- Most of the **secondary education type** professionals take the loan having credit amount range of **[0-1000000]** and they are aged between **40-50**, there are more number of re-payers in both of these cases and loan defaulters are more in **30-40** age category applicants but the number of loan re-payers are more than the number of loan defaulters for all cases.
- Most of the loan applicants who are **married** ones take the loan having credit amount range of **[0-1000000]** and they are aged between **30-50**, in both of these cases there are more number of re-payers and loan defaulters and the number of loan re-payers are more than the number of loan defaulters.
- Most of the loan applicants who **source of income** is through **working** take the loan having credit amount range of **[0-1000000]** and they are aged between **30-50** years in both of these cases there are more number of re-payers and loan defaulters and the number of loan re-payers are more than the number of loan defaulters.

- Most of the loan applicants who are either owning a **house or a apartment** take the loan having credit amount range of **[0-1000000]** and they are aged between **30-50** years, in both of these cases there are more number of re-payers and loan defaulters and the number of loan re-payers are more than the number of loan defaulters.
- Most of the loan applicants having **two family members** with zero children tend to repay the loan on time and those having one children tend to default more if the overall count of client's family is two or three, in both of these cases there are more number of re-payers and loan defaulters and the number of loan re-payers are more than the number of loan defaulters.

Identify Top Correlations for Different Scenarios:

- To find relationship between the two variables, I used Correlation Matrix available in Data Analysis tool from Data Ribbon in Excel by selecting entire numerical column ranges.
- Before finding Correlation, I segmented the dataset using the Target variable into 2 separate parts with Target = 0 that are loan re-payers and Target = 1 that are loan defaulters.



Correlation Matrix for Target = 0 (loan re-payers):

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	FLAG_MOBIL	FLAG_EMP_PHONE	FLAG_WORK_PHONE	FLAG_CONT_MOBILE	FLAG_PHONE	FLAG_EMAIL	CNT_FAM_MEMBERS	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	HOUR_APPR_PROCESS_START	REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	OBS_30_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	DAYS_LAST_PHONE_CHANGE	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY	AMT_REQ_CREDIT_BUREAU_WEEK	AMT_REQ_CREDIT_BUREAU_MON	AMT_REQ_CREDIT_BUREAU_QRT	AMT_REQ_CREDIT_BUREAU_YEAR
CNT_CHILDREN	1.00																																				
AMT_INCOME_TOTAL	0.04	1.00																																			
AMT_CREDIT	0.01	0.38	1.00																																		
AMT_ANNUITY	0.03	0.45	0.77	1.00																																	
AMT_GOODS_PRICE	0.00	0.38	0.99	0.78	1.00																																
DAYS_BIRTH	0.34	0.07	-0.05	0.01	-0.05	1.00																															
DAYS_EMPLOYED	-0.24	-0.16	-0.08	-0.11	-0.08	-0.62	1.00																														
DAYS_REGISTRATION	0.18	0.07	0.01	0.03	0.01	0.34	-0.20	1.00																													
DAYS_ID_PUBLISH	-0.03	0.03	-0.01	0.01	-0.01	-0.01	0.27	-0.27	0.10	1.00																											
FLAG_MOBIL	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	1.00																												
FLAG_EMP_PHONE	0.24	0.16	0.08	0.11	0.07	0.62	-1.00	0.21	0.27	0.00	1.00																										
FLAG_WORK_PHONE	0.05	-0.03	-0.01	-0.02	0.01	0.18	-0.23	0.06	0.05	0.00	0.23	1.00																									
FLAG_CONT_MOBILE	0.00	-0.02	0.02	0.02	-0.01	0.02	0.00	0.00	0.00	-0.02	0.02	1.00																									
FLAG_PHONE	-0.03	0.00	0.02	0.00	0.03	-0.04	0.02	-0.07	-0.03	0.00	-0.02	0.30	0.00	1.00																							
FLAG_EMAIL	0.03	0.09	0.01	0.06	0.01	0.09	-0.07	0.03	0.03	0.00	0.07	-0.01	-0.01	0.02	1.00																						
CNT_FAM_MEMBERS	0.88	0.04	0.06	0.08	0.06	0.28	-0.23	0.17	-0.03	0.00	0.23	0.07	-0.01	-0.02	0.02	1.00																					
REGION_RATING_CLIENT	0.02	-0.21	-0.10	-0.13	-0.10	0.01	0.04	0.08	-0.01	0.00	-0.04	0.00	0.01	-0.09	-0.06	0.02	1.00																				
REGION_RATING_CLIENT_W_CITY	0.02	-0.22	-0.11	-0.14	-0.11	0.01	0.04	0.07	-0.01	0.00	-0.04	0.01	0.01	-0.08	-0.06	-0.06	0.02	0.95	1.00																		
HOUR_APPR_PROCESS_START	-0.01	0.09	0.06	0.05	0.07	0.10	-0.09	0.00	0.04	0.00	0.09	0.03	0.00	0.06	0.02	-0.01	-0.28	-0.26	1.00																		
REG_REGION_NOT_LIVE_REGION	-0.01	0.08	0.03	0.05	0.03	0.06	-0.04	0.03	0.03	0.00	0.04	0.06	0.00	0.01	0.02	-0.01	-0.04	-0.04	0.05	1.00																	
REG_REGION_NOT_WORK_REGION	0.01	0.16	0.06	0.08	0.06	0.10	-0.11	0.03	0.05	0.00	0.11	0.07	0.00	0.00	0.04	0.01	-0.15	-0.14	0.07	0.45	1.00																
LIVE_REGION_NOT_WORK_REGION	0.02	0.15	0.05	0.07	0.05	0.07	-0.10	0.02	0.03	0.00	0.10	0.04	0.00	-0.01	0.04	0.02	-0.15	-0.14	0.06	0.08	0.86	1.00															
REG_CITY_NOT_LIVE_CITY	0.02	0.01	-0.02	-0.01	-0.02	0.18	-0.09	0.07	0.08	0.00	0.09	0.06	-0.04	0.01	0.01	0.04	0.04	0.02	0.34	0.15	0.02	1.00															
REG_CITY_NOT_WORK_CITY	0.07	0.02	-0.01	0.00	-0.01	0.24	-0.25	0.09	0.10	0.00	0.26	0.12	0.00	-0.04	0.01	0.08	0.01	0.03	0.03	0.14	0.24	0.18	0.44	1.00													
LIVE_CITY_NOT_WORK_CITY	0.07	0.02	0.00	0.01	0.00	0.15	-0.22	0.06	0.06	0.00	0.22	0.11	0.00	-0.03	0.00	0.08	-0.02	0.00	0.02	0.00	0.19	0.23	0.03	0.83	1.00												
OBS_30_CNT_SOCIAL_CIRCLE	0.02	-0.03	0.00	-0.01	0.00	0.01	0.01	-0.01	-0.01	0.00	-0.01	-0.02	0.01	-0.03	0.00	0.02	0.04	0.03	-0.01	-0.02	-0.03	-0.02	-0.01	-0.01	0.01	1.00											
DEF_30_CNT_SOCIAL_CIRCLE	0.00	-0.03	-0.01	-0.02	-0.02	0.00	0.02	0.00	0.00	-0.02	-0.01	0.00	-0.03	0.00	0.00	0.01	0.01	0.00	-0.01	-0.01	0.01	0.00	0.00	0.31	1.00												
OBS_60_CNT_SOCIAL_CIRCLE	0.02	-0.03	0.00	-0.01	0.00	0.01	0.01	-0.01	0.00	-0.01	-0.02	0.01	-0.03	0.00	0.02	0.04	0.03	-0.01	-0.02	-0.03	-0.02	-0.01	-0.01	-0.01	0.00	1.00											
DEF_60_CNT_SOCIAL_CIRCLE	0.00	-0.03	-0.02	-0.02	0.00	0.02	0.01	0.00	0.00	-0.02	-0.01	0.00	-0.03	0.00	0.00	0.01	0.01	0.00	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	0.00	0.31	1.00										
DEF_60_CNT_SOCIAL_CIRCLE	0.00	-0.03	-0.02	-0.02	0.00	0.02	0.01	0.00	0.00	-0.02	-0.01	0.00	-0.03	0.00	0.00	0.01	0.01	0.00	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	0.00	0.31	1.00										
DAYS_LAST_PHONE_CHANGE	0.00	-0.05	-0.07	-0.06	-0.07	0.07	0.03	0.05	0.09	-0.01	-0.03	-0.05	-0.03	-0.07	-0.02	-0.03	0.02	0.02	-0.01	0.03	0.04	0.03	0.05	0.04	0.02	-0.01	0.00	-0.02	0.00	1.00							
AMT_REQ_CREDIT_BUREAU_HOUR	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00							
AMT_REQ_CREDIT_BUREAU_DAY	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00							
AMT_REQ_CREDIT_BUREAU_WEEK	0.00	0.01	0.01	0.02	0.01	0.00	-0.01	0.00	0.00	0.00	-0.02	0.00	0.00	-0.02	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00							
AMT_REQ_CREDIT_BUREAU_MON	-0.01	0.07	0.06	0.04	0.07	0.00	-0.03	-0.01	-0.01	0.00	0.03	-0.01	0.01	0.04	0.02	0.00	-0.06	-0.06	0.03	-0.01	0.00	0.01	-0.01	0.00	0.01	0.01	0.00	-0.05	0.01	0.00	-0.01	1.00					
AMT_REQ_CREDIT_BUREAU_QRT	0.00	0.02	0.03	0.01	0.03	-0.02	0.01	0.00	-0.02	0.00	-0.01	0.02	0.00	0.01	0.01	0.00	0.00	-0.01	0.00	-0.01	0.01	0.01	0.01	-0.01	0.01	0.01	-0.01	0.01	0.01	0.01	1.00						
AMT_REQ_CREDIT_BUREAU_YEAR	-0.04	0.03	-0.03	0.00	-0.03	-0.07	0.04	-0.02	-0.04	0.00	-0.04	-0.07	0.02	-0.01	0.04	-0.02	0.01	0.00	-0.03	-0.02	-0.03	-0.02	-0.01	-0.01	0.03	0.02	-0.12	0.00	0.00	0.02	0.02	0.12	1.00				

Top 10 correlation for Target = 0 are:

1. [OBS_60_CNT_SOCIAL_CIRCLE, OBS_30_CNT_SOCIAL_CIRCLE] = 1.00
2. [AMT_CREDIT, AMT_GOODS_PRICE] = 0.99
3. [REGION_RATING_CLIENT, REGION_RATING_CLIENT_W_CITY] = 0.95
4. [CNT_CHILDREN, CNT_FAM_MEMBERS] = 0.88
5. [REG_REGION_NOT_WORK_REGION, LIVE_REGION_NOT_WORK_REGION] = 0.86
6. [DEF_60_CNT_SOCIAL_CIRCLE, DEF_30_CNT_SOCIAL_CIRCLE] = 0.85
7. [REG_CITY_NOT_WORK_CITY, LIVE_CITY_NOT_WORK_CITY] = 0.83
8. [AMT_GOODS_PRICE, AMT_ANNUITY] = 0.78
9. [AMT_ANNUITY, AMT_CREDIT] = 0.77
10. [DAYS_BIRTH, FLAG_EMP_PHONE] = 0.62

Correlation Matrix for Target = 1 (loan defaulters):

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	DAYS_BIRTH	DAYS_EMPLOYED	DAYS_REGISTRATION	DAYS_ID_PUBLISH	FLAG_EMP_PHONE	FLAG_WORK_PHONE	FLAG_CONT_MOBILE	FLAG_PHONE	FLAG_EMAIL	CNT_FAM_MEMBERS	REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	HOUR_APPR_PROCESS_START	REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	REG_CITY_NOT_LIVE_CITY	REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	OBS_30_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	DAYSLAST_PHONE_CHANGE	AMT_REQ_CREDIT_BUREAU_HOUR	AMT_REQ_CREDIT_BUREAU_DAY	AMT_REQ_CREDIT_BUREAU_WEEK	AMT_REQ_CREDIT_BUREAU_MON	AMT_REQ_CREDIT_BUREAU_QRT	AMT_REQ_CREDIT_BUREAU_YEAR
CNT_CHILDREN	1.00																																		
AMT_INCOME_TOTAL	0.01	1.00																																	
AMT_CREDIT	0.01	0.02	1.00																																
AMT_ANNUITY	0.03	0.02	0.75	1.00																															
AMT_GOODS_PRICE	0.00	0.01	0.98	0.75	1.00																														
DAYS_BIRTH	0.25	0.01	-0.14	-0.01	-0.14	1.00																													
DAYS_EMPLOYED	-0.19	-0.01	0.02	-0.08	0.02	-0.58	1.00																												
DAYS_REGISTRATION	0.15	-0.01	-0.04	0.02	-0.04	0.29	-0.19	1.00																											
DAYS_ID_PUBLISH	-0.04	-0.01	-0.04	-0.02	-0.05	0.25	-0.23	0.09	1.00																										
FLAG_EMP_PHONE	0.19	0.01	-0.02	0.08	-0.02	0.58	-1.00	0.19	0.23	1.00																									
FLAG_WORK_PHONE	0.06	-0.01	-0.05	-0.04	-0.03	0.17	-0.21	0.06	0.04	0.21	1.00																								
FLAG_CONT_MOBILE	0.01	0.00	0.03	0.03	0.03	-0.01	-0.01	0.01	0.01	0.02	1.00																								
FLAG_PHONE	-0.01	-0.01	0.04	0.00	0.05	-0.05	0.03	-0.07	-0.04	-0.03	0.30	1.00																							
FLAG_EMAIL	0.01	0.00	0.00	0.10	0.00	0.08	-0.05	-0.01	0.03	0.05	0.02	0.01	1.00																						
CNT_FAM_MEMBERS	0.89	0.01	0.06	0.08	0.05	0.20	-0.18	0.15	-0.04	0.18	0.07	0.00	0.00	1.00																					
REGION_RATING_CLIENT	0.06	-0.01	-0.05	-0.06	-0.05	0.05	-0.01	0.12	0.03	0.01	0.02	0.03	-0.06	-0.02	0.06	1.00																			
REGION_RATING_CLIENT_W_CITY	0.05	-0.01	-0.05	-0.08	-0.06	0.04	0.00	0.11	0.01	0.00	0.03	0.03	-0.06	-0.02	0.06	0.95	1.00																		
HOUR_APPR_PROCESS_START	-0.01	0.01	0.05	0.04	0.06	0.06	-0.05	-0.06	0.01	0.05	0.05	-0.01	0.06	-0.01	-0.02	-0.28	-0.25	1.00																	
REG_REGION_NOT_LIVE_REGION	-0.02	0.00	0.01	0.03	0.01	0.04	-0.04	0.02	0.02	0.04	0.07	-0.05	0.01	0.02	0.00	-0.03	-0.03	0.05	1.00																
REG_REGION_NOT_WORK_REGION	-0.01	0.00	0.02	0.07	0.02	0.08	-0.08	0.02	0.04	0.09	0.10	0.01	0.02	0.02	-0.01	-0.10	-0.10	0.08	0.53	1.00															
LIVE_REGION_NOT_WORK_REGION	0.00	0.00	0.03	0.07	0.07	0.04	-0.07	0.01	0.03	0.07	0.07	0.01	0.03	0.01	-0.01	-0.12	-0.12	0.07	0.10	0.81	1.00														
REG_CITY_NOT_LIVE_CITY	0.00	-0.01	-0.05	-0.02	-0.05	0.15	-0.09	0.06	0.06	0.09	0.02	-0.01	-0.04	0.00	0.01	0.05	0.05	0.01	0.34	0.18	0.03	1.00													
REG_CITY_NOT_WORK_CITY	0.05	-0.01	-0.04	0.00	-0.04	0.23	-0.25	0.10	0.08	0.25	0.10	0.02	-0.01	0.00	0.05	0.02	0.04	0.00	0.15	0.23	0.16	0.47	1.00												
LIVE_CITY_NOT_WORK_CITY	0.06	-0.01	-0.01	0.01	-0.01	0.14	-0.20	0.07	0.04	0.20	0.10	0.02	0.02	-0.01	0.01	-0.01	0.00	0.17	0.22	-0.02	0.78	1.00													
OBS_30_CNT_SOCIAL_CIRCLE	0.02	-0.01	0.03	0.01	0.03	-0.01	0.00	-0.01	-0.03	0.00	-0.07	0.01	-0.05	-0.02	0.04	0.03	0.02	-0.02	-0.03	-0.02	-0.05	-0.04	-0.02	1.00											
DEF_30_CNT_SOCIAL_CIRCLE	-0.01	-0.01	-0.03	-0.03	-0.02	-0.02	0.03	0.00	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.01	0.02	0.01	0.02	0.01	0.00	-0.02	-0.03	0.37	1.00											
OBS_60_CNT_SOCIAL_CIRCLE	0.02	-0.01	0.03	0.01	0.03	-0.01	0.00	-0.01	-0.03	0.00	-0.07	0.01	-0.05	-0.02	0.04	0.03	0.02	-0.02	-0.03	-0.02	-0.05	-0.04	-0.02	1.00	0.37	1.00									
DEF_60_CNT_SOCIAL_CIRCLE	-0.02	-0.01	-0.03	-0.04	-0.02	-0.03	0.02	-0.01	-0.03	-0.02	-0.03	-0.04	-0.02	-0.03	-0.01	0.00	0.00	0.00	0.00	-0.02	0.30	0.89	0.30	1.00											
DAYSLAST_PHONE_CHANGE	0.01	0.01	-0.12	-0.10	-0.13	0.12	-0.02	0.08	0.14	0.02	0.00	-0.01	-0.07	0.02	-0.01	0.03	0.02	-0.04	0.02	0.02	0.01	0.07	0.07	0.04	-0.02	0.00	-0.02	0.02	1.00						
AMT_REQ_CREDIT_BUREAU_HOUR	0.00	0.00	0.02	0.04	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.02	0.00	-0.01	-0.01	-0.03	-0.01	0.02	0.03	0.00	0.02	0.01	-0.01	-0.01	0.01	0.00	-0.01	1.00						
AMT_REQ_CREDIT_BUREAU_DAY	-0.03	0.00	-0.01	-0.02	-0.01	-0.02	0.05	0.00	-0.01	-0.05	0.00	0.00	0.02	0.00	-0.03	0.02	0.02	0.00	0.00	0.01	0.01	-0.02	-0.01	-0.01	-0.01	0.35	1.00								
AMT_REQ_CREDIT_BUREAU_WEEK	-0.03	0.00	0.00	0.03	0.00	-0.01	0.02	-0.02	0.02	-0.02	-0.02	0.01	-0.02	0.03	-0.03	0.00	0.00	-0.01	-0.07	0.07	0.05	-0.02	0.04	-0.04	-0.04	-0.02	0.02	0.01	0.02	-0.06	0.00	-0.02	0.00		
AMT_REQ_CREDIT_BUREAU_MON	0.01	0.00	0.08	0.07	0.08	-0.01	-0.03	0.00	-0.04	0.03	-0.03	0.01	0.00	0.02	0.02	-0.07	-0.07	0.07	0.05	-0.02	0.04	-0.04	-0.04	-0.02	0.02	0.01	0.02	0.01	-0.06	0.00	-0.02	0.00	1.00		
AMT_REQ_CREDIT_BUREAU_QRT	-0.01	0.00	-0.02	0.00	-0.02	-0.01	0.02	-0.01	-0.03	-0.02	-0.07	0.00	-0.03	0.03	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.05	-0.04	0.03	0.02	0.04	0.03	0.00	0.03	0.01	0.02	1.00	
AMT_REQ_CREDIT_BUREAU_YEAR	-0.03	-0.01	-0.02	0.00	-0.02	-0.09	0.02	-0.03	-0.08	-0.02	-0.09	0.03	-0.03	0.05	-0.01	0.01	0.01	-0.04	-0.03	-0.03	-0.01	-0.02	-0.03	-0.01	0.05	0.02	0.05	0.02	-0.10	0.01	0.03	0.03	0.04	0.10	1.00

Top 10 correlation for Target = 1 are:

1. [OBS_60_CNT_SOCIAL_CIRCLE, OBS_30_CNT_SOCIAL_CIRCLE] = 1.00
2. [AMT_CREDIT, AMT_GOODS_PRICE] = 0.98
3. [REGION_RATING_CLIENT, REGION_RATING_CLIENT_W_CITY] = 0.95
4. [CNT_CHILDREN, CNT_FAM_MEMBERS] = 0.89
5. [DEF_60_CNT_SOCIAL_CIRCLE, DEF_30_CNT_SOCIAL_CIRCLE] = 0.89
6. [REG_REGION_NOT_WORK_REGION, LIVE_REGION_NOT_WORK_REGION] = 0.81
7. [REG_CITY_NOT_WORK_CITY, LIVE_CITY_NOT_WORK_CITY] = 0.78
8. [AMT_CREDIT, AMT_ANNUITY] = 0.75
9. [AMT_ANNUITY, AMT_GOODS_PRICE] = 0.75
10. [DAYS_BIRTH, FLAG_EMP_PHONE] = 0.58

Results and Inferences:

- It clearly indicates that in most cases the more the applicants repay the loans in same segment the more they tend to default in the same segment.
- In most of cases, it shows that the number of loan defaulters are less than the number of loan re-payers. Banks should find out how they can further more reduce the number of loan defaulters.
- For HOUSING_TYPE segment, banks should also focus on gaining customers from other apartment type like Rented Apartment, Co-op apartment, Municipal apartment, Office apartment on how they can get more loan re-payers in this categories without failing to default.
- For EDUCATION_TYPE segment, banks should also focus on gaining customers from other education types like academic degree, incomplete higher, lower secondary categories on how they can get more loan re-payers in this categories without failing to default.

- For FAMILY_STATUS segment, banks should also focus on gaining customers from other family statuses like civil marriage, widow, separate, single kind of people on how they can get more loan re-payers in this categories without failing to default.
- For INCOME_TYPE segment, banks should also focus on gaining customers from other income source types like businessman, student, unemployed people on how they can get more loan re-payers in this segments without failing to default.
- Banks should also focus on clients having higher number of family members more than 3 on how they can get more loan takers and ensuring they can pay the loan on time.
- Banks should find out how they can offer loan at higher credit amount ranges and also find out how they can attract more customers having higher income ranges.

Conclusion:

- This project successfully brings out the required insights. All the questions were answered by carrying out analysis using Excel pivot tables and pivotcharts.
- The insights drawn were important and it answered the required questions. For much deep-dive understanding and other findings, you can refer to this [application.xlsx](#).

